DOCUMENT RESUME

ED 301 598 TM 012 590

AUTHOR Love, Gayle A.
TITLE The Myth of the Correction for Guessing Formulas with Feasible Alternatives Suggested.
PUB DATE 87
NOTE 19p.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Guessing (Tests); *Scoring Formulas; Self Esteem; Teacher Made Tests; Test Construction; *Test Wiseness
IDENTIFIERS *Correction for Guessing; *Test Instructions

ABSTRACT

In a review of relevant literature, it is argued that correction for guessing formulas should not be used. It is contended that such formulas correct for guessing that does not really exist in a noticeable amount, penalize those students who have low self-esteem and self-confidence, correct for errors that are not necessarily errors, benefit risk takers, lower validity, and disregard the benefits of partial knowledge. An improved strategy would be to teach students test-wiseness skills, eliminating a source of error, so that students would know how to take tests correctly. More specifically, students need to learn test-wiseness skills so they can budget their time, avoid making unnecessary errors, correctly use their partial knowledge to their best advantage, and identify test-constructor errors; thus, eliminating the bias placed on those students who do not pick up these test-taking skills naturally. To this end, teachers must be taught test-wiseness skills, improve their tests and test instructions, and learn which errors to avoid when writing their tests. (Author/TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED301598

The Myth of the Correction for Guessing Formulas with Feasible Alternatives Suggested

Gayle Love

Eastern New Mexico University

Psychology Department

Portales, New Mexico

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

GAYLE A. LOVE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM012 590

BEST COPY AVAILABLE

Abstract

Correction for guessing formulas should not be used. They correct for guessing that does not really exist in a noticable amount and penalize those students that are low in self-esteem and confidence. They also correct for errors that are not necessarily errors. A far better strategy would be to teach students test-wiseness skills, eliminating a source of error there, so that the students will know how to correctly take tests. To do this, one must also teach teachers test-wiseness skills and help them to improve their tests and instructions.

The Myth of the Correction for Guessing Formulas with Feasible Alternatives Suggested

How to mark tests has been a question that has plagued teachers and test-givers everywhere since the first test was constructed and administered. Then, also, came the problem of how to grade fairly and equally without favoring any one person.

Correction for guessing formulas are a method test-givers use to attempt to correct test grades of their examinees to nullify the effects of guessing. Several people (Albanese, 1986; Lord, 1975) have argued that applying a correction formula is the only fair way to administer a test, but there has been much controversy over the use of correction formulas both pro and con. If a teacher really wishes to score a test impartially and fairly, however, they should refrain from using correction for guessing formulas that penalize students and instead teach the skills their students need to learn to take a test and give them properly constructed tests.

Correction formulas of varying degrees of complexity have been developed. "All are variations of the original method of subtracting the number wrong from the number right, ignoring omitted items, and considering the remainder as the significant measure" (Sheriffs & Boomer, 1954, p. 81). Usually a fraction equal to one over the number of distractors (one in the case of a true-false item) is multiplied times the number wrong, giving a formula similar to the following:

$$\#\text{Right} - \left(\frac{1}{\#\text{distractors}} \times \#\text{wrong}\right) = \text{Corrected Score}$$

This formula is supposed to rule out an answer the test-taker might have gotten

right simply by guessing and the laws of probability. The alternative to using the correction for guessing formula is to use the standard Number Right formula, which is simply the number of correctly marked answers on the test.

There is a major problem with using a correction for guessing formula in grading tests. "Formula scoring is based on the assumption that examinees either know the answer to a test item or else choose among the alternative responses at random" (Lord, 1975, p. 7). Those who express concern over the effects of guessing often seem to assume that what takes place in the examination setting is blind guessing (Mehrens & Lehmann, 1986, p 337), but "guessing on multiple choice tests is not the serious problem it has often been thought to be" (p. 342).

The explicit assumption made by those in favor of the correction for guessing formula is that, when appropriate instructions are provided, "the only difference between an answer sheet obtained under formula scoring directions would be that omitted responses would replace random guessing" (Cross & Frary, 1977, p. 313). "It is further assumed that all the wrong answers plus some of the right ones are chance answers or guesses" (Little, 1962, p. 245).

Many studies have been conducted using the effects of guessing as their subject matter. "The theory that students tend to guess correctly a large number of items was not supported..." (Jackson, 1955, p. 78). In fact, according to Mehrens and Lehmann (1986), "luck in guessing may raise or lower a particular student's score by a mark or two, but it will not make any noticeable difference unless the test is unreasonably short" (p. 339). That seems reasonable. If a test has five questions, guessing one correctly would change the grade by 20%, but if the test had one hundred questions, guessing one

correctly would only change the grade by 1%.

It has also been shown that examinees rarely guess blindly (Sarnacki, 1979, p. 257). "If Lois consistently 'guesses' more successfully than Jim, we will soon have to accept the conclusion that Lois is not guessing - at least not blindly - and that Lois really does know something that Jim does not" (Mehrens & Lehmann, 1986, p. 339). Should she be penalized for this? "Students ... rarely find themselves in situations where blind guessing is all they can do" (p. 337). "Typically, examinees have some partial information about an item" (Lord, 1975, p. 7).

It has even been suggested that "'do-not-guess' instructions should not be given, even when formula scoring is used" (Cureton, 1966, p. 44). If students are encouraged through test directions and scoring to use their partial knowledge in making informed guesses, "those with higher levels of partial knowledge will be correct more often than those with lower levels" (Mehrens & Lehmann, 1986, p. 338) of partial knowledge. Examinees who make informed guesses do better than those who do not guess, since those who make guesses are operating on the basis of partial knowledge, and "partial information is more likely than a random guess to lead to a correct response" (Angoff & Schrader, 1986, p. 242).

Studies (Cross & Frary, 1977; Bliss, 1980) have been made where students were asked to "guess" on answers they had omitted under correction formula scoring directions. It was found that "examinees who omitted items under formula scoring directions were omitting items for which they would, when forced to guess, on the average, earn significantly more points than would be expected by chance" (Cross & Frary, 1977, p. 320). Not only that, but "the

consistency with which this test-taking behavior has been found across ages and the findings that formula scoring may especially tend to penalize more able students raises an ethical question about the advisability of using formula scoring directions" (Bliss, 1980, p. 152). Mehrens and Lehmann (1986) have concluded that "validity can be enhanced by informed guessing" (p. 337).

The personality factor of the examinees must also enter into consideration when discussing correction for guessing formula. The "ability to 'figure the odds', subjective levels of confidence in own judgement, and willing to gamble or to take a chance, are among the variables which might be expected to influence the right-minus-wrong measure of achievement" (Sheriffs & Boomer, 1954, p. 82). One can see how "certain students either can not or will not comply with directions appropriate for formula scoring" (Cross & Frary, 1977, p. 313). In fact, "even with the application of a correction, those who obey the 'do-not-guess' instructions are disadvantaged and those who disregard the instructions profit from doing so" (Mehrens & Lehmann, 1986, p. 34). Students with "high levels of introversion, anxiety, and low self-esteem were penalized by the use of right-minus-wrong scoring" while a significant difference was not found in the number-right control group (Sheriffs & Boomer, 1954, p. 86). It seems that when formula scoring directions are given "those students who are most intimidated by this situation will omit questions that they could and should have answered, and their scores will suffer as a result" (Mehrens & Lehmann, 1986, p. 341). In fact, "many investigators have found that this method [correction for guessing] penalizes more cautious examinees" (Jaradat & Sawaged, 1986, p. 369). It has been concluded that "the most honest advice to students would be that they should answer every question

on the test, even if the instructions tell them otherwise" (Mehrens & Lehmann, 1986, p. 341)!

A further problem with the assumption that all wrong answers and some correct answers on tests are the product of guessing is that "some wrong answers are not marked in accordance with the laws of chance, but instead are based on positive misinformation" (Little, 1962, p. 248). Positive misinformation is a situation where the examinee has mislearned the information and had honestly believed themselves to be marking the correct answer. The item writer attempts to make each wrong answer so plausible that every examinee who does not posses the desired skill or ability will select a wrong response. "In actual practice, this aim of the item writer is not fully realized, but it is doubtlessly often sufficiently realized that the standard [correction] formula markedly overcorrects" (Traxler, 1951, p. 349). Sax and Collet (1986) believed that "correction formulas should not be used because they treat mistakes produced through misinformation as though they occurred by chance..." (p. 1127).

The problem of over-correction is a serious one. When the total number of wrong answers are used in the correction formula "they yield a correction factor which overcorrects and gives the student a lower estimate than is warranted by his true knowledge and ability" (Little, 1962, p. 247). "If a person answers a question wrongly in the genuine belief that it is correct, not only does he fail to score on that question, but he loses a fraction of a mark from somewhere else - a penalty for a 'guess' he did not make" (Mehrens & Lehmann, 1986, p. 342)! This is particularly severe in the case of a true-false test, where "the 'correction' for a wrong answer is enough to cancel out a

previously correct answer" (p. 341).

The correction formula does not always "correct" in the way it is supposed to correct, either. "The guessing correction is applied to all wrong answers whether they are to easy or [to] hard items. No account is taken of the difference between right answers to hard questions and right answers to easy ones" (Smith, 1986, p. 361). One would expect misinformation to have a greater effect on these harder items. Further, " a pattern of wrong answers 'corrected for guessing' and hence reduced in score could signify the presence of silly errors which needed amendment by score increase to assure an accurate measure" (p. 361). Little (1962) promotes the idea that "the basic assumption underlying the correction for guessing is misleading and may make it appear that the examinee knows the answers to fewer questions than he really does know" (p. 248). Generally it has been shown that "the application of correction formulae results in a wider range of test scores and in a downward adjustment of average scores" (Sheriffs & Boomer, 1954, p. 81), but the "correction formulas do not alter the rank assigned to the students significantly" (Sax & Collet, 1968, p. 1127).

So, if the use of correction for guessing formulas are going to penalize high anxiety and low self-esteem students while benefiting risk takers, lower validity, disregard the benefits of partial knowledge, and correct for blind guessing that is not really occurring without significantly altering rank - only lowering the average score, what alternatives are there? How can differences among students be minimized so tests measure what they are supposed to measure - content knowledge? The answer is a simple one. Instruction is needed: instruction for students in test-wiseness skills and instruction for

teachers in proper test construction techniques.

Mentioning test-wiseness also seems to create controversy. Test-wiseness is "a subject's capacity to utilize the characteristics and format of the test and/or the test-taking situation to receive a high score" (Millman, Bishop, & Ebel, 1965, p. 707). Further, test-wiseness "suggests a cognitive ability that one may employ on a variety of tests, regardless of the nature of the tests' subjective content" (Sarnacki, 1979, p. 253). This may imply cheating or an unfair advantage for some students, but test-wiseness skills are naturally acquired by many. In an unstructured questionnaire where students were asked to help a new student with taking tests, many test-wiseness skills were suggested (Millman et al., 1965, p. 709) - though they were not identified as such.

It may be useful at this point to explain exactly what is meant by test-wiseness. Millman et al. (1965) gives the most comprehensive listing of test-wiseness principles. He divides the skills into two groups. One group consists of time- using strategies, strategies to avoid mechanical errors, guessing strategies to maximize the use of partial knowledge, and deductive reasoning strategies (p. 711). The second group is dependant upon the test-constructor and focuses on intent consideration strategies (what kind of answer should it be) and cue using strategies to exploit test writing errors made by the test-constructor (p. 712-3). "Most of the principles [of test-wiseness] enunciated were applicable to objective tests, although test-wiseness is probably just as important a factor in relation to essay tests (for example, how to plan an essay, how to write from an outline, time usage, how to 'skirt around' an area of weakness...)" (Mehrens & Lehmann, 1986, p. 436). Since "there seems to be little possibility of completely removing the effects of test-wiseness, perhaps

more thought should be given to the teaching of test-wiseness to those students low in test-wiseness skills" (Brueggeman, 1987, p. 162).

Besides, as Mehrens and Lehmann pointed out, "it is neither possible nor desirable to take test-wiseness skills away from those who already have them" (1986, p. 338). In fact, "training in test-wiseness could reduce individual differences on it [tests]" (Fagely, 1986, p. 76) and "more error in measurements is likely to originate from students who have too little, rather than too much skill in taking tests" (Scruggs, Lifson, & Steven, 1985, p. 343). Brueggemann (1987) reminds us that studies show that, if no instruction is given, test-wiseness is a stable condition over time and that "students that are penalized on test scores because of low test-wiseness tend to be penalized ... every time they take a test" (p. 162).

What can be done about this? "The problematic effects of test-wiseness [or a lack of test-wiseness] are best ameliorated through test-wiseness training" (Sarnacki, 1979, p. 267). Obviously, poor comprehenders may benefit from more knowledge of test-taking strategies" (Ritter & Idol-Maestas, 1986, p. 356). In fact, in several studies conducted, teaching test-wiseness skills was found to be "beneficial to all experimental students (both high and low intelligence)" (p. 356) and "the ability to pick up test-wiseness cues may be demonstrated at all levels of intelligence ..." (Sarnacki, 1979, p. 269). Wilson (1986) goes as far to claim that "acquiring appropriate study and test-taking skills and attitudes is an important development task that students must master if they are to make satisfactory academic progress" (p. 297). Brueggemann (1987) continues in the same vein by suggesting that "deficiencies in one or more of the abilities prerequisite to taking a test [test-wiseness] will attenuate the results, thereby

depreciating the test effectiveness to discriminate validly only those variables it was designed to measure" (p. 159).

Mehrens and Lehmann (1986) believe it is "both possible and desirable to teach test-wiseness to those who lack it" (p. 336). Sarnacki (1979) showed that a number of methods may be used to teach numerous aspects of test-wiseness (p. 276) and that poor comprehenders can be taught to use test-taking skills within a heterogeneous group (Ritter & Idol-Maestas, 1986, p. 356). Unfortunately it was found that "none of the programs have focused on one major component of study skills instruction - preparing for and taking examinations [test-wiseness]" (Wilson, 1986, p. 297). In fact, "instruction in test-taking is not commonly included in school curricula and instruction" (Ritter & Idol-Maestas, 1986, p. 350).

"Test-wiseness is just one aspect of the topic of testing; however, it is one that classroom teachers can help students with" (Brueggemann, 1987, p. 159). Wilson (1986) points out how "without special assistance, students may come to view tests and examinations as negative experiences and continue to be handicapped by inappropriate test-taking skills and attitudes throughout their educational careers" (p. 297). "Further, the test-wiseness program conveys a considerable informational content and is not unlike placebo controls often used in anxiety research" (Bajtelsmit, 1977, p. 340). "Teachers should attempt to teach all of their students to be test-wise" (Mehrens & Lehmann, 1986, p. 347).

While many important and useful strategies are discussed, a distressing thought should be that nearly two-thirds of the "Outline of Test-wiseness Principles" deals with how students should learn to profit from a test-constructors errors (Millman et al., 1965, p. 713). It seems that

"test-wiseness conveys a greater advantage when the test is ... poorly constructed" (Mehrens & Lehmann, 1986, p. 336). Although not all test-wiseness skills capitalize upon a poorly constructed test, time using strategies are entirely independent of the test-constructor, "the deductive test-taker is able to make use of various item faults to improve his or her test score" (Sarnacki, 1979, p. 259).

"Teacher-made tests frequently exhibit test-wiseness cues ... their test-writing skills are not as precise [as professional test writers], so item flaws inevitably occur" (Sarnacki, 1979, p. 265). To compound the problem, "as compared to the professional test constructor, most teachers are relatively naive of test-wiseness principles" (p. 266). Since "test-wiseness matters more on a poorly constructed test than on a well-constructed test, a teacher who has developed adequate skills in test-construction is part-way to solving the problem" (Mehrens & Lehmann, 1986, p. 336). Another approach to eliminating cues on tests suggests instructing test-makers in the principles of both proper test construction in general, and test-wiseness in particular (Sarnacki, 1979, p. 274). "Careful attention to important details such as grammatical consistency, logic, and absence of lengths for correct versus incorrect alternatives in construction of plausible distractors" (Kolstad, Kolstad, & Wagner, 1985-6, p. 44). "Almost all textbooks in educational measurement provide rules intended to help the test constructor avoid giving away the answer" (Millman et al., 1965, p. 708) and to avoid confusing the examinee.

It seems both teachers and their students could benefit from this teacher training. "On a well constructed exam ... few secondary clues are available, thereby eliminating one type of test-wiseness [secondary cue-using] as a source

of variance" (Bajtelsmit, 1977, p. 240). A test-writer can refine or improve an item by incorporating better ideas into the item or by changing the dictation in the statement of the ideas (Dressel & Schmid, 1953, p. 579). After all,

> writing a good multiple choice item is not merely the perfunctory job of juxtaposing a correct answer with several false or misleading answers picked out of the air, but is a complex and difficult problem which careful selection and relating of alternative ideas to a worthwhile concept so that the student has an articulated and meaningful task to perform (p. 574).

That is a very complicated task, and student test grades are a highly valuable commodity to risk. Surely teachers should have the proper training before being sent out to literally write thousands of these questions. It is no wonder that "certain idiosyncracies of the test-constructor may manifest themselves in various flaws in item construction" (Sarnacki, 1979, p. 259), or that biases occur in keying correct responses - over 42% of test-constructors have a tendency to place the correct answer at the center of the response distribution (p. 280). Sarnacki (1979) has also shown that "poor sentence structure and difficult vocabulary may cause test-takers to focus on these shortcomings instead of attending to subject matter" (p. 264) and that "knowledge of the test-constructors idiosyncracies in test making and grading will aid the examinee in maximizing scores ..." (p. 267). Faults were found to make the items easier (McMorris, Brown, Snyder, & Pruzek, 1972, p. 287) and "flawed test items allow the introduction of a source of variance other than item content" (Sarnacki, 1979, p. 263). However, if the test-constructor is cognizant of test-wiseness skills "they may be employed as bait, intended to lure the

individual into selecting an incorrect alternative" (p. 263) and insuring that questions are read carefully for content, also.

Lastly, teachers should "be cognizant of the type of instructions" (Bauer, 1973, p. 30) used. In some cases the complaint has been registered, with reason, that "understanding of the directions is more difficult than working the test items" (Dressel & Schmid, 1953, p. 574). If formula scoring directions were modified to read "guess at none, answer only those you can use partial information on" (Lord, 1975, p. 8) then students would not be so heavily penalized, and they would know when to omit answers. "Examinees often omit items they could score better than chance; but often their instructions were inadequate for the present purpose (p. 8). Their instructions briefly indicated or described the scoring method, but "they gave the examinee not guidance to help him find the proper strategy for responding.... The difficulty can be corrected by giving better test instructions" (p. 9).

So what exactly needs to be corrected? The correction for guessing formula needs to have its use discontinued. "The benefits of removing a number of sources of bias variation outweigh the problems arising from random responding" (Bauer, 1973, p. 32) which are very low. "The weight of evidence is now fairly clearly against their use" (Mehrens & Lehmann, 1986, p. 340).

Instead, time should be spent to properly educate our students and teachers. Students need to learn how to effectively take tests. They need to learn test-wiseness skills so they can budget their time, avoid making unnecessary errors, correctly use their partial knowledge to their best advantage, and identify test-constructor errors, eliminating the bias placed on those students who do not pick up these test-taking skills naturally. Teachers

also need to be trained in these areas. They need to know the skills so that they can teach them to the students, and they need to know what errors to avoid making when writing their tests. If teachers and students are taught how to properly write and take tests, guessing will not be necessary and the correction for guessing formula will be useless.

References

Albanese, M. A. (1986). The correction for guessing: A further analysis of Angoff and Schrader. Journal of Educational Measurement, 23, 225-35.

Angoff, W. H. & Schrader, W. B. (1986). 'A rejoinder to Albanese:' the correction for guessing: A further analysis of Angoff and Schrader. Journal of Educational Measurement, 23(3), 237-43.

Arnold, J. C. & Arnold, P. L. (1970). On scoring multiple choice exams allowing for partial knowledge. Journal for Experimental Education, 39, 8-13.

Bajtelsmit, J. (1977). Test wiseness and systematic desensitization programs for increasing test-taking skill. Journal of Educational Measurement, 14, 335-342.

Bauer, D. H. (1973). Error sources in aptitude and achievement test scores: A review and recommendation. Measurements in Evaluation and Guidance, 6, 28-34.

Bliss, L. B. (1980). A test of Lord's assumption regarding examinee guessing behavior on multiple choice tests using elementary school students. Journal of Educational Measurement, 17, 147-152.

Borrello, G. M. & Thompson, B. (1985). Correlates of selected test-wise skills. Journal of Experimental Education, 53(3), 124-8.

Brueggemann, L. V. (1987). What teachers know about test-wiseness. Reading Horizons, 27(3), 159-63.

Cross, L. H. & Frary, R. B. (1977). An empirical test of Lord's theoretical results regarding formula scoring of multiple choice tests. Journal of Educational Measurement, 14, 313-321.

Cureton, E. E. (1966). The correction for guessing. Journal of Experimental Education, 4, 44-7.

Dressel, P. L. & Schmid, J. (1953). Some modification of the multiple choice item. Educational and Psychological Measurement, 13, 574-95.

Fagley, N. S. (1987). Positional response bias in multiple choice tests of learning: Its relation to test-wiseness and guessing strategy. Journal of Educational Psychology, 79, 95-7.

Jackson, R. A. (1955). Guessing and test performance. Educational and Psychological Measurement, 15, 74-9.

Jaradat, D. & Sawaged, S. (1986). The subset selection technique for multiple choice tests: an empirical inquiry. Journal of Educational Measurement, 23, 369-76.

Kolstad, R. K. and others. (1983). The failure of distractors on complex multiple choice items to prevent guessing. Educational Research Quarterly, 8(2), 44-50.

Kolstad, R. K., Kolstad, R. A., & Wagner, M. J. (1985-6). Performance on 3 choice vs. 5 choice multiple choice items that measure different skills. Educational Research Quarterly, 10(2), 4-8.

Little, E. B. (1962). Overcorrection for guessing on multiple choice test scoring. Journal of Educational Research, 55(6), 245-52.

Lord, F. M. (1975). Formula scoring and numbers-right scoring. Journal of Educational Measurement, 12, 7-11.

McMorris, R. F., Brown, J. A., Snyder, G. W. & Pruzek, R. M. (1972). Effects of violating item construction principles. Journal of Educational Measurement, 9, 287-95.

Mehrens, W. A. & Lehmann, I. J. (1986). Measurements and evaluations in education and psychology. New York: Holt, Rinehart, and Winston.

Millman, J., Bishop, C. H. & Ebel, R. (1965). An analysis of test-wiseness. Educational and Psychological Measurement, 25, 707-26.

Ritter, S. & Idol-Maestas, L. (1986). Teaching middle school students to use a test-taking strategy. Journal of Educational Research, 79(6), 350-7.

Sarnacki, R. E. (1979). An examination of test-wiseness in the cognitive test domain. Review of Educational Research, 49, 252-279.

Sax, G. & Collet, L. S. (1968). An analysis and empirical study of the effect of guessing formulas on reliability and validity. Educational and Psychological Measurement, 29, 1127-34.

Scruggs, T. E., Lifson, A. & Steven, A. (1985). Current conceptions of

test-wiseness: myths and realities. School Psychological Review, 14(3), 339-50.

Sheriffs, A. C. & Boomer, D. S. (1954). Who is penalized by the penalty for guessing? Journal of Educational Psychology, 45, 81-90.

Smith, R. M. (1986). Person fit in the Rasch model. Educational and Psychological Measurement, 45, 359-77.

Traxler, A. E. (1951). Administering and scoring the objective test. In E. F. Lindquist (Ed.), Educational Measurement (pp. 349). Washington: American Council on Education.

Wilson, N. S. (1986). Preparing for examinations: A classroom guidance unit. School Counselor, 33(4), 297-305.